

**PROCEEDINGS OF
THE 2013 INTERNATIONAL CONFERENCE ON
SCIENTIFIC COMPUTING**

CSC 2013

Editors

**Hamid R. Arabnia
George A. Gravvanis
George Jandieri
Ashu M. G. Solo
Fernando G. Tinetti**



WORLDCOMP'13

July 22-25, 2013

Las Vegas Nevada, USA

www.world-academy-of-science.org

©CSREA Press

This volume contains papers presented at The 2013 International Conference on Scientific Computing (CSC'13). Their inclusion in this publication does not necessarily constitute endorsements by editors or by the publisher.

Copyright and Reprint Permission

Copying without a fee is permitted provided that the copies are not made or distributed for direct commercial advantage, and credit to source is given. Abstracting is permitted with credit to the source. Please contact the publisher for other copying, reprint, or republication permission.

Copyright © 2013 CSREA Press

ISBN: 1-60132-238-0

Printed in the United States of America

CSREA Press

U. S. A.

A combined algorithm for analyzing structural controllability and observability of complex networks

Luis Úbeda^{1,2}, Carlos Herrera^{1,2,3}, Iker Barriales^{1,2}, Pedro J. Zufiria^{1,2}, and Mariluz Congosto⁴

¹Depto. Matemática Aplicada a las Tecnologías de la Información, ETSI Telecomunicación, Universidad Politécnica de Madrid (UPM), Spain

²Cátedra Orange. ETSI Telecomunicación, Universidad Politécnica de Madrid (UPM), Spain

³Depto. Civil and Environmental Engineering. Massachusetts Institute of Technology, USA

⁴Departamento de Telemática. Universidad Carlos III de Madrid, Spain

E-mail: lubeme23@gmail.com, chyague@gmail.com, ibarriales@ono.com, pedro.zufiria@upm.es, mariluz.congosto@gmail.com

Abstract—In this paper a combined algorithm for analyzing structural controllability and observability of complex networks is presented. The algorithm addresses the two fundamental properties to guarantee structural controllability of a system: the absence of dilations and the accessibility of all nodes. The first problem is reformulated as a Maximum Matching search and it is addressed via the Hopcroft-Karp algorithm; the second problem is solved via a new wiring algorithm. Both algorithms can be combined to efficiently determine the number of required controllers and observers as well as the new required connections in order to guarantee controllability and observability in real complex networks. An application to a Twitter social network with over 100,000 nodes illustrates the proposed algorithms.

Keywords: Complex networks, controllability, observability, maximum matching, Twitter

1. Introduction

Classical dynamical system theory is based on the characterization of systems via time-evolution models (e.g. difference or differential equations) [18], [24]. Such characterization can be employed to perform two fundamental and dual tasks in dynamical system theory: on the one hand, the estimation of system's internal state from the measurement of accessible system outputs, defining the so-called observation problem; on the other hand, the modification of system's internal state via the injection of appropriate system inputs, defining the so-called control problem [6]. These problems have been thoroughly studied in the literature so that observability and controllability conditions are well defined, specially for linear systems [3], [8].

The study of complex systems composed by many interconnected elements can be addressed by analyzing the

underlying network (or graph) which characterizes these interconnections [1]. Such network gathers some fundamental structural system properties which in general do not depend on the specific system parameter values. In general, both node dynamics and their connectivity are relevant in complex systems control [4], [5], [22]; here in this paper we focus on structural aspects to provide a fundamental reference information for a detailed controllability analysis.

Along this line, structural controllability theory has been developed [14], [15], [20], [21] in order to characterize the implications of system internal connectivity in its state space controllability. The existing results can be formalized via either matrix algebraic structural properties or graph theoretical properties. Parallel results have also been obtained for the dual structural observability problem [16], [23]. Supported by this duality, the exposition of results in this paper will mainly refer to the controllability problem.

System structural controllability becomes determined by both its internal connection structure and the set of inputs defined in such system. Given the network underlying the system connectivity structure and a set of inputs, a system is structurally controllable if and only if all nodes (or vertices) are accessible from the inputs and the network does not have dilations [14].

The first part of this paper addresses the problem of guaranteeing system structural controllability by making use of a minimum number of input controllers. Graph theoretical tools are mainly employed to find such minimum number of control inputs for avoiding dilations, where the concept of Matching becomes crucial [15], [21]. In addition, the concept of wiring is developed to guarantee accessibility of all nodes preserving such minimum number of control inputs.

The second part of the paper illustrates the application of the proposed algorithms to both the controllability and the (dual) observability problems.

In the following section we summarize some of the basic results concerning system structural controllability.

Corresponding author: Pedro J. Zufiria. Depto. Matemática Aplicada a las Tecnologías de la Información, ETSI Telecomunicación, Universidad Politécnica de Madrid (UPM), Ciudad Universitaria s/n, E-28040 Madrid, Spain.

2. Structural controllability fundamental results

As stated in Section 1, a network is structurally controllable if and only if all nodes are accessible from the inputs and the network has no dilation [15]. We start addressing the analysis of dilations.

2.1 Dilation and Matching

A graph (or network) is defined by $G := (V, E)$, a pair of sets where V is the set of vertices (or nodes) and E the set of edges (or links). Every edge belonging to E is defined as a pair of vertices $(v_i, v_j) \in E$ where $v_i, v_j \in V$. For directed graphs, the pair defining an edge is ordered.

A network has a dilation if we can find a subset $S \subset V$ of the vertices of the graph that verifies:

$$|S| > |T(S)| \quad (1)$$

where $T(S)$ represents the set of vertices that point to any vertex $v_i \in S$.

On the other hand, a bipartite graph is defined as $G_b := (X, Y, E_b)$ where X and Y are two (potentially) different sets, and E_b is the set of edges linking one element of X with an element of Y , i.e. $E_b = \{(v_i^X, v_j^Y) : v_i^X \in X, v_j^Y \in Y\}$. Given a set of nodes $S \subseteq Y$, we define $T(S) \subseteq X$ such that $T(S) = \{v_i^X \in X : \exists (v_i^X, v_j^Y) \text{ with } v_j^Y \in S\}$.

Graph Theory provides the following result:

Theorem 1 (Hall): A bipartite graph $G_b := (X, Y, E_b)$ has a matching which covers every vertex in X if and only if

$$|T(S)| \geq |S|, \forall S \subseteq Y \quad (2)$$

Hall's theorem provides a criterion for bipartite graphs to have a perfect matching, as long as they also verify:

$$|X| = |Y| \quad (3)$$

This condition on the existence of a perfect matching is condensed in the following corollary:

Corollary 1.1: A bipartite graph $G_b := (X, Y, E_b)$ has a perfect matching if and only if it verifies both (2) and (3).

Given a graph $G := (V, E)$ one can define its associated bipartite graph $G_b := (V^+, V^-, E_b)$ where $V^+ = V^- = V$ and E_b is constructed as follows

$$E_b = \{(v_i^+, v_j^-) : \exists (v_i, v_j) \in E\}$$

Defined this way, $G_b : (V^+, V^-, E_b)$ will obviously verify (3). Therefore, a network has no dilation if and only if it has a perfect matching.

Alternatively, based on the pioneering work of [14], the Maximum Matching (MM) Algorithm is proposed in [15] as a good tool to determine the minimum number of inputs required to guarantee that there are no dilations in a network. Precisely, the number of required inputs would correspond with the number of vertices not matched by an edge in the maximum matching of the network.

2.2 Accessibility and wiring

However, as stated earlier, the matching criterion verifies the condition of no dilation in the network, but not the accessibility condition. The network might present some internal structures that being matched by edges of the maximum matching might not be accessible from the inputs. This happens with matching solutions containing cycles. In these cases, it is necessary to add new connection edges from the calculated inputs to the non-accessible structures; we call this process *wiring*. Note that this process keeps the number of control inputs unchanged.

In the following Section, a version of the MM Algorithm is proposed and analyzed. In addition, a new wiring algorithm is developed, in order to find the structures matched by the MM that are not accessible from the inputs, and to create the required external edges from the inputs to a node of such structures.

3. The combined MM and wiring algorithm

3.1 The Hopcroft-Karp algorithm

Complex networks are usually formed by a very large amount of nodes (or vertices) and links (or edges). Hence, the time efficiency of the Maximum Matching (MM) detection algorithm is very important when dealing with this type of huge networks. For this purpose, the Hopcroft-Karp algorithm has been selected which runs in $O(\sqrt{V}E)$ time [11].

The Hopcroft-Karp algorithm directly deals with non-directed bipartite graphs; nevertheless it can also be applied to general non-bipartite directed graphs, provided an appropriate transformation is previously performed. As mentioned earlier, given the original graph $G := (V, E)$ one can define its associated bipartite graph $G_b := (V^+, V^-, E_b)$ where $V^+ = V^- = V$. Note that the edges in the bipartite graph adjacent to any $v_i^+ \in V^+$ represent the out-links of $v_i \in V$ and the edges adjacent to $v_i^- \in V^-$ represent the in-links of v_i .

Given a matching M_b of G_b , if $(v_i^+, v_j^-) \in M_b$, then $(v_i, v_j) \in M$ where M is the corresponding matching of G . $(v_i^+, v_j^-) \in M_b$ makes nodes v_i^+, v_j^- to be matched on the non-directed bipartite graph G_b while $(v_i, v_j) \in M$ makes v_j to be matched on the directed graph G . Therefore, once we compute the Hopcroft-Karp algorithm over G_b and obtain a maximum matching M_b^* (where $*$ stands for maximum), the unmatched nodes of the set V^- will be the unmatched nodes of the matching M^* in G .

Following the minimum inputs theorem [15] the unmatched nodes given a maximum matching will be the driver nodes of the network, so a controller node must be linked to each one of these driver nodes.

3.2 The accessibility and wiring algorithms

As stated in section 2.2 it is not enough to obtain the maximum matching M^* of the graph G to find the nodes that need to be controlled. Placing the controllers on the unmatched nodes only guarantees the absence of dilations on the graph; hence, it is necessary to additionally perform a search for the inaccessible nodes from the controllers of the network. From section 2.2 it follows that these inaccessible nodes will be placed in the loops of the considered maximum matching M^* of G . The method proposed here consists then on looking for loops on M^* in order to determine those which are not accessible from any controller node. As we are dealing with a directed network, looking for loops in the network given by the matching is equivalent to looking for its strongly connected components [19]; thus, it is possible to perform this search by applying the Tarjan algorithm which runs in $O(|V| + |E|)$ time [13]. Since we are only interested in the strongly connected components of the sub-graph given by M^* (in which $|E| = |M^*|$ and $|V| = |M^*| - 1$ or $|V| = |M^*|$ in the case of a perfect matching, where $|M^*|$ stands for the number of links in the maximum matching), Tarjan's algorithm will run in $O(|M^*|)$. Since the computational cost of this search is much lower than the Hopcroft-Karp maximum matching search over the whole graph, the running time of the Tarjan algorithm is almost negligible.

Once the loops of M^* have been found, an accessibility analysis from the controller nodes must be done. A possible strategy to accomplish this purpose is making a BFS (Breadth-First Search [13] for every found loop. Nevertheless, the BFS algorithm must be modified in order to allow the root and the goal of the search to be sets of nodes instead of single nodes, the root set being the nodes of the loop under analysis and the goal set being the unmatched nodes of the network. In addition, an early termination mechanism has also been implemented since once a controller node is found, accessibility is guaranteed and therefore the search can be finished. Based on these modifications, the original BFS algorithm running time of $O(|V| + |E|)$ for each loop can be significantly reduced. Note also that this modified BFS algorithm will deal with short depth paths when dealing with complex networks. The small world [25] property of the network guarantees that if the loop is accessible it will likely be at a short distance from a controller node. Furthermore, it has also been proved that in this kind of networks a giant component arises that contains most of the nodes in the network; hence, if the loop is not accessible from any controller node it will probably belong to a small component, resulting also in a very short depth search.

Finally, a wiring process for the discovered inaccessible loops is performed, adding links between a node of each of these loops and a controller node. Putting all the steps together, the search of the nodes that need to be controlled on a graph G can be summarized in the following four steps:

- 1) Find the maximum matching M^* of G and link a new different controller node to each unmatched node.
- 2) Find the loops in M^* .
- 3) Find those loops of M^* that are not accessible from the controller nodes obtained at step 1.
- 4) For each found inaccessible loop select a node belonging to it and wire it up to a controller node.

4. Controllability and observability on a real complex network

During the latest years the study of social networks using digital communication records as a proxy for human interactions has enormously grown. Phone calls, texts, or online communications allow researchers to analyze human interaction in unprecedented scales. In order to provide a real world example of controllability in complex networks we have considered the interaction between users in the social site Twitter. First, the data collection process is described and some basic characteristics of the social network are provided. Finally, a controllability (and observability) analysis using the techniques explained in previous sections is performed.

4.1 Data description and network characteristics

Our data-set consists of over 11 million tweets containing interactions between users in the Madrid urban area during February 2012. Interactions can be either mentions (user A mentions B in a message) or re-tweets (user A broadcasts to its followers a message from B). In both cases, the interaction is modeled as a directed link from B to A. This interaction network has been referred previously in the literature as the *dynamic* network as opposed to the *static* network where links represent a declared follower-followee relationship. Data have been gathered using Twitter's Streaming API, which allows for geographic filtering; users need to have authored at least one tweet in the data-set to be considered as a node in the resulting network (i.e., re-tweets not authored in Madrid were discarded).

The resulting network consists of 119,217 nodes, whose degrees distributions are presented in figure 1. Average degree in the network is $\langle k_{in} \rangle = \langle k_{out} \rangle = 7.83$. Although both distributions present a certain long tail, this tail is longer for the out degree one. Precisely the maximum out degree is 10,150 while the maximum in degree is 663. The explanation for this difference lies on the limited cognitive capacity of human beings: a person may influence millions of people, but an individual cannot be influenced by the behaviors of millions of people. This reasoning is supported by empirical evidence in primates [7] and more recently in Twitter itself [10]. Apart from an heterogeneous degree, the network exhibits small-world behavior [25]: it presents a high number of triangles (clustering coefficient 0.015, two orders of magnitude more than the equivalent random graph)

and short diameter (estimated $\langle l \rangle = 4.14 \pm 0.23$ using 100 random different origins to traverse the graph). Additionally, 21.01% of the relationships are mutual and 96.63% of the nodes belong to a giant connected component. Overall, this network presents general characteristics which are similar to the ones observed in other complex networks [17].

4.2 Analysis of results

Applying the algorithms described on section 3 to the dynamic Twitter network we obtained a total of $N_D = 27.129$ required controller nodes. This set of nodes is sufficient to avoid dilations on the network; however in order to guarantee full structural controllability all nodes must be accessible; thus a search for inaccessible nodes is performed according to the wiring algorithm described in Section 2.2. An average of 800 wired nodes (with a standard deviation two orders of magnitude smaller) are required for the network to satisfy the accessibility condition, and therefore to be controllable. From the above analysis one can see that this network needs almost 25% of the nodes to be directly influenced by a driver node; this means that it cannot be easily controlled.

Since the maximum matching of a network is not unique, a Montecarlo analysis was performed in order to quantify the importance of every node on the different control configurations of the network. The algorithm described on section 3 may provide a different set of controlled nodes for each iteration; since we are interested in estimating the frequency each node shows up in the controlled nodes set, such algorithm has been run up to 40,000 times to guarantee a low estimation variance. Gathering nodes with the same frequency, Figure 2 is obtained. It can be seen that almost all the nodes have a well defined role on the control configuration, since they are either always ($p = 0$) or never ($p = 1$) present in the controlled nodes set. An interesting fact can also be observed: the histogram shows high peaks for values of $p = \frac{1}{4}$, $p = \frac{1}{3}$, $p = \frac{1}{2}$, $p = \frac{2}{3}$, $p = \frac{3}{4}$, $p = \frac{4}{5}$... This phenomenon was explained in [23] to be a consequence of the effect of some repetitive microstructures (motifs) that commonly appear on these networks.

A parallel study was made regarding the observability of the network [23], obtaining similar results to the above commented for controllability. However a significantly bigger amount of nodes was required on the wiring stage to satisfy the accessibility condition. This difference comes from the asymmetry between the in-degree and out-degree distributions of the network.

In order to assess the information obtained by the controllability and observability analysis, a cross-correlation analysis of both indices with some usual metrics used in complex networks was performed. Such selected metrics are

the PageRank [2], betweenness centrality¹, out/in-degree, and a physics based influence index proposed in [9]. Figure 3 displays the results of this analysis, where one can observe that both controllability and observability are almost totally uncorrelated with all the considered metrics; this means that the information provided by this analysis can complementarily shed some new light in the node characterization of complex networks.

5. Concluding remarks

An algorithm for analyzing structural controllability and observability of complex networks has been presented. The algorithm combines a Maximum Matching search and a new wiring algorithm to efficiently determine the number of required controllers and observers as well as the new required connections in the network. The application to a Twitter social network with over 100,000 nodes has illustrated the applicability of the algorithm on a real complex network. The results suggest that these measures (number of required controllers and observers) provide new metrics to characterize the network structural properties.

Acknowledgments

The authors want to acknowledge the financial support of Orange (Spain and France), in the framework of Cátedra Orange at the ETSI Telecomunicación in the Universidad Politécnica de Madrid (UPM), Spain. The work has been also partially supported by projects MTM2010-15102 of Ministerio de Ciencia e Innovación, and Q10 0930-144 of UPM, Spain.

References

- [1] A. L. Barabási and R. Albert, "Emergence of scaling in random networks" *Science*, 286, pp. 509-512, 1999.
- [2] Brin, S. and Page, L., "The anatomy of a large-scale hypertextual Web search engine", *Computer networks and ISDN systems*, 30, 107-117, 1998.
- [3] T.-C. Chen, "Linear System Theory: the state space approach", Oxford University Press, 1999.
- [4] S. P. Cornelius, W. L. Kath and A. E. Motter, "Controlling Complex Networks with Compensatory Perturbations", arXiv.1105.3726, 2011.
- [5] N. J. Cowan, E. J. Chastain, D. A. Vilhena, J. S. Freudenberg and C. T. Bergson, "Nodal dynamics, not degree distributions, determine the structural controllability of complex networks", arXiv.1106.2573, 2011.
- [6] R. C. Dorf and R. H. Bishop, "Modern Control Systems", Pearson, 2011.
- [7] Dunbar, R. I. "Neocortex size as a constraint on group size in primates". *Journal of Human Evolution*, 22(6), 469-493, 1992.
- [8] B. Friedland, "Control System Design: An introduction to state-space methods", Dover, 1986.

¹Betweenness centrality is a common topological measure in complex networks, defined as the fraction of shortest paths in the network which cross a certain node: classic example is the Anchorage airport in Alaska, which is not a *hub* (does not have many connections to others), but it is indeed a crucial node in the air traffic network since allows refuel for many America-Asia connections.

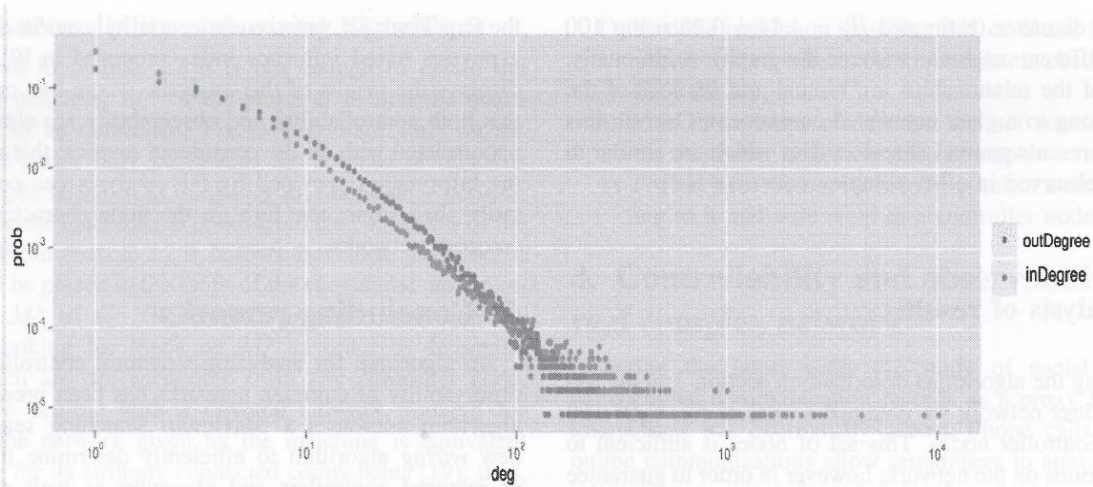


Fig. 1: In-degree and out-degree distributions of the network.

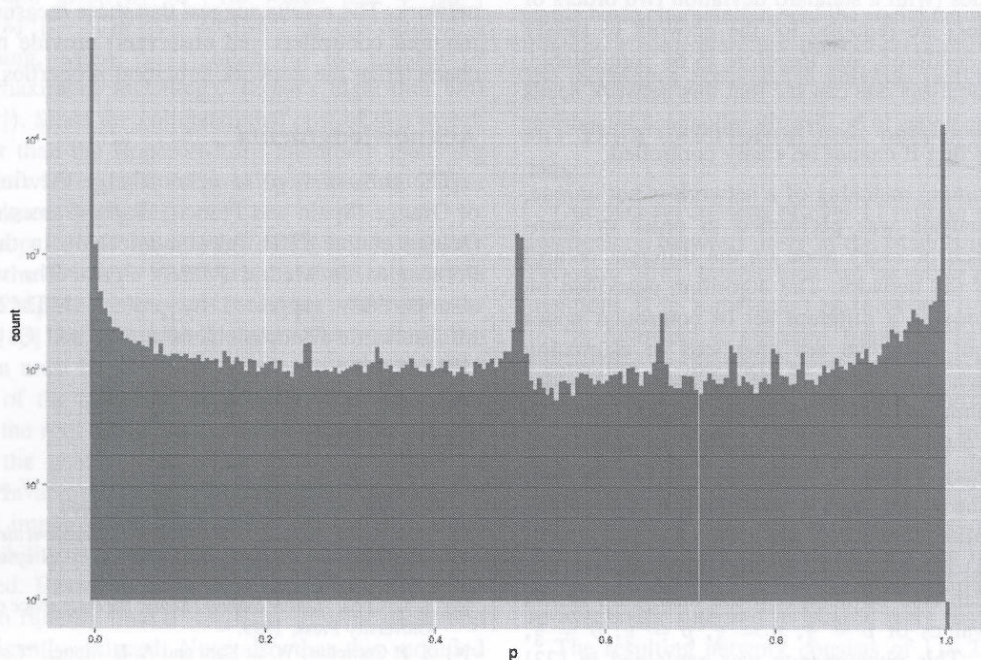


Fig. 2: Frequency of appearance of the nodes on the controller nodes set.

- [9] D. Gayo-Avello, D. J. Brenes, D. Fernández-Fernández, M. E. Fernández-Menéndez and R. García-Suárez, "De retibus socialibus et legibus momenti" (On social networks and the laws of influence, EPL (Europhysics Letters), (94) 38001, doi: 10.1209/0295-5075/94/38001.
- [10] Goncalves, B., Perra, N., and Vespignani, A. "Validation of Dunbar's number in Twitter conversations". arXiv preprint arXiv:1105.5170.,2011.
- [11] J. E. Hopcroft and R. M. Karp, "An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs", SIAM J. Comput. 2, 225 (1973).
- [12] R. E. Kalman, J. Soc. Indus. and Appl. Math. Ser. A 1, 152, 1963.
- [13] Knuth, Donald E., "The Art Of Computer Programming Vol 1. 3rd ed.", Boston: Addison-Wesley, ISBN 0-201-89683-4, 1997.
- [14] Ching-Tai Lin, "Structural Controllability", IEEE Trans. AC-19(3) 201-208, 1974.
- [15] Y. Liu, J. Slotine and A. L. Barabási, "Controllability of Complex Networks", Nature 473, 168, 2011.
- [16] Y. Liu, J. Slotine and A. L. Barabási, "Observability of Complex Networks", PNAS (DOI: 10.1073/1215508110), 2013.
- [17] Newman, M. E. J., "The structure and function of complex networks", SIAM review, 45 (2) 167-256 , 2003.
- [18] B. Friedland, "Introduction to Dynamic Systems: Theory, Models and Applications", John Wiley & Sons, 1979.
- [19] Sedgewick, R., "Graph algorithms", Algorithms, Addison-Wesley, ISBN 0-201-06672-6, 1983.
- [20] R. W. Shields and J. B. Pearson, "Structural Controllability of Multi-Input Linear Systems", IEEE Transactions on Automatic Control,

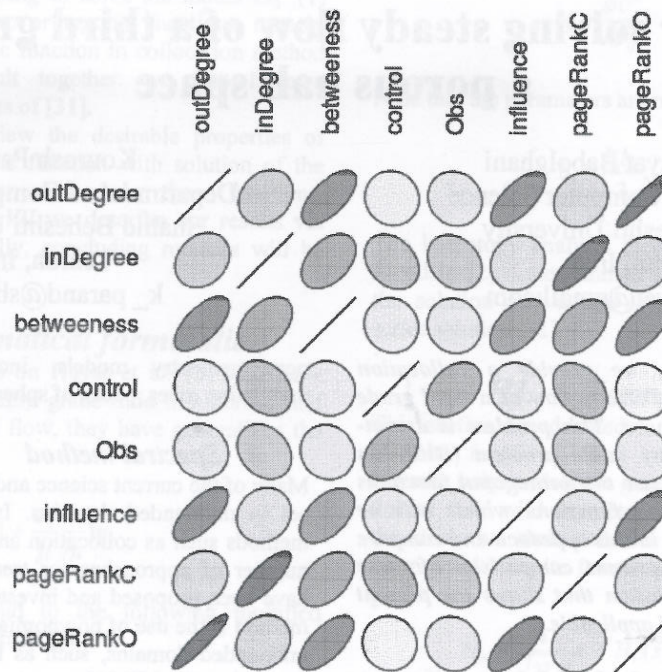


Fig. 3: Cross-correlations between different network metrics.

VOL. AC-21, NO. 2, APRIL 1976.

- [21] R. W. Shields and J. B. Pearson, "Author's Reply to Corrections to Structural Controllability of Multi-Input Linear Systems", IEEE Transactions on Automatic Control, VOL. AC-23, NO. 3, JUNE 1978.
- [22] J. Sun, S. P. Cornelius, W. L. Kath and A. E. Motter "Comment on 'Controllability of Complex Networks with Nonlinear Dynamics'", arXiv.1108.5739, 2011.
- [23] L. A. Úbeda Medina, "Controlabilidad y observabilidad en redes complejas", Proyecto Fin de Carrera, ETSI Telecomunicación, Universidad Politécnica de Madrid, Octubre 2012.
- [24] M. Vidyasagar, "Nonlinear Systems Analysis", SIAM, 2002.
- [25] Watts, D. J. and Strogatz, S. H., "Collective dynamics of small-world networks", Nature, 393, 6684, 440-442, 1998.